

A Composition-Transferable Machine Learning Potential for LiCl-KCl Molten Salts Validated by HEXRD

Jicheng Guo

*Chemical and Fuel Cycle Technologies Division,
Argonne National Laboratory, Lemont, IL 60439, USA*

Logan Ward and Yadu Babuji

Data Science and Learning Division, Argonne National Laboratory, Lemont, IL 60439, USA

Nathaniel Hoyt and Mark Williamson

Chemical and Fuel Cycle Technologies Division, Argonne National Laboratory, Lemont, IL 60439, USA

Ian Foster

Data Science and Learning Division, Argonne National Laboratory, Lemont, IL 60439, USA

Nicholas Jackson

Department of Chemistry, University of Illinois, Urbana-Champaign, Urbana, IL 61801, USA

Chris Benmore

X-ray Science Division, Argonne National Laboratory, Lemont, IL 60439, USA

Ganesh Sivaraman*

*Data Science and Learning Division,
Argonne National Laboratory, Lemont, IL 60439, USA*

(Dated: July 12, 2022)

Unraveling the liquid structure of multi-component molten salts is challenging due to the difficulty in conducting and interpreting high temperature diffraction experiments. Motivated by this challenge, we developed composition-transferable Gaussian Approximation Potentials (GAP) for molten LiCl-KCl. A DFT-SCAN accurate GAP is active learned from only ~ 1100 training configurations drawn from 10 unique mixture compositions enriched with metadynamics. The GAP-computed structures show strong agreement across HEXRD experiments, including for an eutectic not explicitly included in model training, thereby opening the possibility for composition discovery.

I. INTRODUCTION

Molten salts (MS) are a class of high temperature ionic liquids relevant to liquid metal batteries, concentrated solar power systems, and molten salt reactors [1–5]. Critical to technological applications of MS are the eutectic mixtures of alkali/alkali-earth halides, the melting temperatures of which can be lowered by tuning the mixture composition [6, 7]. Tuning the thermophysical properties of multi-component salts requires a precise atomistic understanding of the liquid structure; challenges for this task include (a) the difficulty of *in situ* experimental measurements due to the high reactivity of molten salts with moisture and oxygen, (b) lack of *a priori* knowledge regarding optimal eutectic compositions for arbitrary MS chemistries [8], and (c) insufficiency of simple, empirical interatomic potentials for capturing the complex many-body interactions, such as polarizability, in-

herent to molten salt [9–14].

Machine learning (ML) has enabled [15–18] a new generation of low-cost interatomic potentials (IP) that provide access to quantum mechanically accurate many-body potential energy surfaces (PES) for condensed phases [19–28]. ML-IP’s based on neural networks and kernel methods can construct complex, high-dimensional functional forms with large parameter sets (i.e. 10^4 - 10^5) by learning from *ab initio* datasets [29]. These ML-IP can drive simulations of atomic processes with *ab initio* accuracy that bypass the length scale limitations imposed by traditional *ab initio* methods.

Transferability is the ability of an ML-IP to predict, with consistent accuracy, properties of configurations that differ significantly from the ones used for training (e.g. composition change) [30]. Recent efforts have begun applying these methods to single and multi-component MS [31–35]. However, current ML-IP for MS are generally fit to a small number of mixture compositions (including *a priori* known eutectic compositions), with neural network based ML-IP’s needing $O(10^5)$ training samples per composition [36], and lack transferability to

* gsivaraman@anl.gov

arbitrary or unknown mixture compositions. Classical empirical potential suffer less from these limitations due to their well defined functional forms that ensure some modicum of transferability. This lack of transferability limits application of existing ML-IP to simple interpolative efforts for MS compositions used in model training and inhibits their application to eutectic discovery for arbitrary MS chemistries. Thus, there is an urgent need for ML-IP of MS exhibiting multi-component, compositional transferability across the space of possible anion and cation combinations.

LiCl-KCl mixtures exhibit high potential for molten salt reactors, pyroprocessing and energy storage applications due to their ideal physiochemical properties such as low melting point, high solubility of fission products, and high heat capacity [37–39]. In this study, we utilize LiCl-KCl as a model system to showcase the development of a ML-IP with broad compositional transferability, and apply it to ascertain the composition-dependent structure of LiCl-KCl mixtures. The compositional transferability of this methodology opens the door for *a priori* prediction, by molecular modeling, of eutectic compositions for arbitrary MS. Finally, we demonstrated the use of ML-IP to accurately compute thermal conductivity of the eutectic molten LiCl-KCl.

II. METHODOLOGY

A. A combined workflow for GAP modeling fitting and experimental validation

We have designed a combined experimental and modeling workflow for investigating multiple compositions of LiCl-KCl melts (Fig. 1). The modeling component of the workflow leads to the generation of a multi-composition ML-IP, specifically a Gaussian approximation potential (GAP) [19], for LiCl-KCl. The GAP model uses two-body squared exponential and many-body smooth overlap of atomic positions (SOAP) kernel functions [40, 41] to measure chemical similarity between local chemical neighborhoods. The short-range GAP model is a reasonable choice given that the goal is to model bulk liquid structure [31, 42]. However, explicit inclusion of long-range electrostatics is necessary for modeling scenarios requiring inclusion of nonisotropic chemical environments such as interfaces or applied electric fields [43]. GAP model training utilizes atomic configurations drawn from a diverse set of melt compositions. Configurational sampling (Fig. 1a) was initialized with 10 unique compositions hand picked by the experimentalist, listed here as molar fractions of KCl (100%, 90%, 80%, 67%, 58%, 50%, 33.3%, 30%, 20%, and 10%). Initial simulation densities are fixed to literature values [44]. At each composition, a small unit cell of 60–64 atoms is used to sample a range of melted atomic configurations by using the Born-Mayer-Huggins-Tosi-Fumi (BMHTF) rigid ion model (RIM) at an elevated temperature of 2100K. Configuration sam-

pling was explicitly *not performed* for the middle of the composition map between 50% to 33.3% KCl molar fraction. This conscious choice was made in order to not include the actual eutectic composition of 58.5-41.5 mol% LiCl-KCl as part of the training set. By doing so, the ability of the generated GAP model to accurately predict compositions that are not explicitly included in the training data can be assessed.

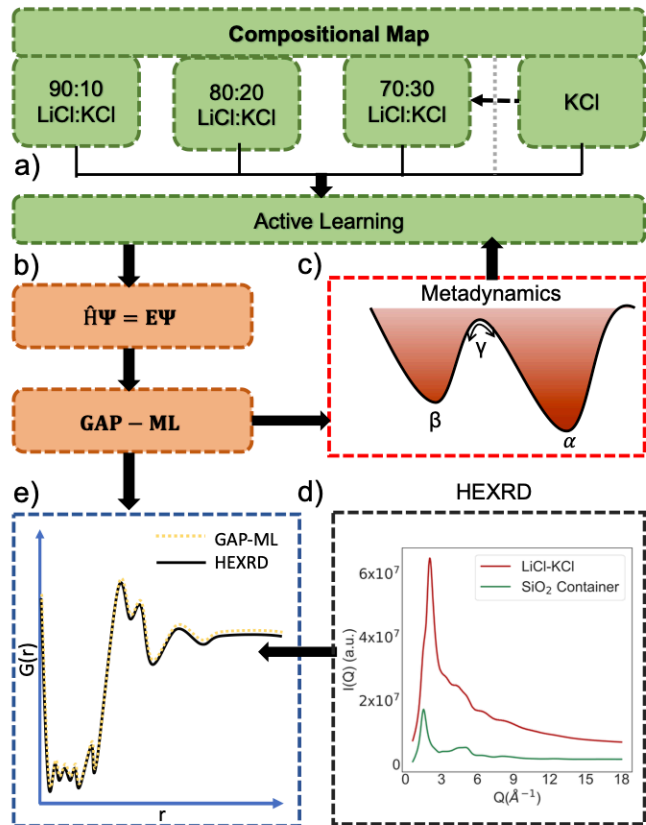


FIG. 1. Workflow for mapping and validating multi-composition LiCl-KCl melts. (a) Sample the configuration space for 10 unique compositions of pure KCl to 10 mol% KCl mixture with LiCl. Each composition is melted by using the rigid ion model at 2100K. Clustering-based AL enables down selection. Ensemble simulations and AL are orchestrated by using the Colmena framework [45]. (b) Perform single point DFT for the AL samples and fit the GAP-ML model. (c) Enrich the configuration space by using metadynamics [46] on the GAP-ML based MD. (d) High energy X-ray diffraction experiments. Illustration of X-ray intensity measurement. (e) Perform rigorous validation of GAP-ML driven MD simulation by using the high energy X-ray pair distribution functions (PDFs).

Configurational sampling for each composition uses approximately 20,000 melt configurations drawn at 2100K and processed using active learning (AL). The AL approach consists of an unsupervised clustering algorithm combined with Bayesian optimization for on-the-fly hyperparameter tuning of the GAP [47, 48]. A detailed

description of the AL sampling strategy builds up on our earlier dedicated study on LiCl melts [42], is reproduced in Appendix A. Single point DFT calculations (Fig. 1b) are performed on the AL-extracted configurations by using the SCAN exchange correlation (XC) functional [49], which shows superior performance compared to generalized gradient approximation (GGA) XC functionals [50, 51]. DFT single point calculations are performed using the Vienna ab initio simulation package [52]. The SCAN exchange-correlation functional and projector-augmented wave method are employed [49, 53]. A large plane wave cutoff of 700 eV with an electronic convergence criterion of 10^{-7} eV is used. A Γ -centered $1 \times 1 \times 1$ k-mesh is used for reciprocal sampling.

An initial GAP model is fitted to the AL-extracted configurations for the 10 compositions (Fig. 1b), with “labels” computed using DFT-SCAN calculations. Since all training configurations are drawn from equilibrium MD simulations, there is no guarantee of configurational or compositional transferability for the GAP model. To circumvent the limitations of Boltzmann sampling, strategies such as random structure search and enhanced sampling have been applied to enrich training databases [54, 55]. Here, we employ metadynamics to construct a history-dependent repulsive potential as a function of a set of collective variables (CVs), providing access to a large configuration space of ion pair coordination environments that would be unexplored by simple equilibrium MD [46, 56]. We hypothesize that the increased configurational diversity at each training composition should manifest in improved compositional transferability across the LiCl-KCl composition range.

Enhanced sampling is performed by using the well-tempered variation of metadynamics for an equal fraction of LiCl and KCl (i.e., 50%) [57]. A system size of 64 atoms near the melting temperature (~ 757 K) is simulated by using the atomic pair coordination collective variable (CV) parameterized by the first minimum of the pair distribution function (PDF). The metadynamics calculations are performed using the plumed 2 packages [58]. A coordination number collective variable is chosen. For any two arbitrary chemical species groups A and B, they are estimated using the following switch function:

$$C_{AB} = \sum_{i \in A, j \in B} \frac{1 - \left(\frac{r_{ij}}{r_0}\right)^6}{1 - \left(\frac{r_{ij}}{r_0}\right)^{12}} \quad (1)$$

with r_{ij} being the distances between the atom-atom pairs from the species groups.

A system size of 64 size with 16 anion-cation pairs each for Li-Cl and K-Cl is used for sampling. The r_0 was parametrized from the first minima of the partial pair distributions functions. For Li-Cl and K-Cl the r_0 of 3.35 Å and 4.25 Å respectively are found to be optimal. The metadynamics used an initial Gaussian height of 40 kJ/mol. The Gaussian widths of 1.45, 2.58 are used for C_{LiCl} , and C_{KCl} respectively. A Gaussian was deposited every 250 fs with a bias factor equal

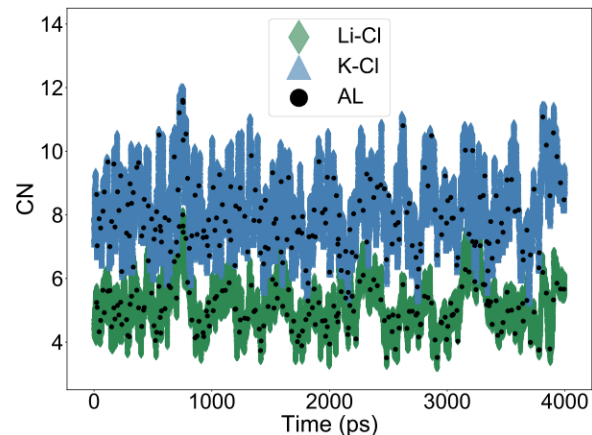


FIG. 2. Evolution of the anion-cation pair collective variable for the 50:50 composition over 4 ns of well-tempered metadynamics performed using the intermediate GAP-ML model. The configurations selected by active learning for single point DFT are shown as black dots. The y-axis shows coordination number (CN) averaged over 16 anion-cation pair used in the CV definition.

to 50. The evolution of the CV over 4 ns is shown in Fig 2. In equilibrium MD the sampled configuration will be dominated by ~ 4 and ~ 7 coordinated Li-Cl and K-Cl, respectively. Metadynamics facilitates efficient exploration of out-of-equilibrium regions of configuration space. The configurations obtained from over 4 ns of metadynamics simulation were passed to AL and down-selected. It is observed that AL efficiently samples from near and far away from equilibrium regions of metadynamics configurations, thereby enriching the training database (Fig. 2). The final GAP model is trained by using the metadynamics-enriched dataset. The details of the GAP training database and hyperparameters are listed in Table I and Table II. The entire GAP model is generated with only 1,127 training samples which is also the total number of single point DFT performed as a part of this study, effectively bypassing the need for expensive *ab initio* molecular dynamics for training dataset generation which would be a Herculean task at a DFT-SCAN level of theory. A single compute node benchmark with 64 atoms unit cell of 50-50 mol% showed $15,000\times$ speedup of GAP-MD (~ 0.042 s/MD time step) relative to DFT-SCAN (~ 618 s/MD time step).

System	Number of samples
Melt	928
Metadynamics	196
Isolated atoms	3
Total	1127

TABLE I. GAP training database

The GAP model (Table I, II) validated on 120 independent test samples drawn at the 10 composition showed an root mean square error in energy and force

Parameter Name	Two-body	SOAP
Cut off ($\text{\AA}$)	5.92	5.92
Sparse method	Uniform	CUR
Sparse points	65	1200
Delta (eV)	2.74	0.78
(l_{max}, n_{max})	-	(4,8)

TABLE II. GAP model hyperparameter

of 5 meV/atom and $0.12 (\pm 0.04)$ eV/ $\text{\AA}$, respectively. The RMSE in energy can be attributed to the target accuracy set for the AL sampling strategy as described in Appendix A.

B. High energy X-ray diffraction experiments

High energy X-ray diffraction (HEXRD) experiments were performed to characterize the structure of molten LiCl-KCl mixtures in parallel with the modeling work (Fig. 1). For HEXRD experiments, LiCl-KCl mixture samples were prepared in an ultra-high purity Ar glove-box (<1 ppm O_2 , <1 ppm H_2O). Samples with the desired LiCl-KCl ratio were melted in glassy carbon crucibles to create homogeneous mixtures, then the solidified mixtures were crushed and loaded into silica glass (SiO_2) ampules and sealed under vacuum for the experiments. HEXRD experiments were performed at beamline 6-ID-D of the Advanced Photon Source, Argonne National Laboratory. LiCl-KCl mixtures were heated to the desired temperatures at which diffraction data was recorded. The diffraction patterns of empty furnace and empty glass ampules were also recorded for background subtraction. The diffraction data of LiCl-KCl with several compositions including 70-30 mol%, 58.5-41.5 mol% (eutectic composition) and 50-50 mol% were recorded at multiple temperatures. Temperature dependence is studied by measuring the structures of 70-30 mol% at $\sim 822\text{K}$, 797K , 772K , and 747K . The 58.5-41.5 mol% (eutectic composition) and 50-50 mol% are measured near the melting temperatures of $\sim 645\text{K}$ and 757K , respectively.

III. RESULTS

A. Validation of GAP model for composition transferability

To verify the composition transferability of the GAP model, validation is performed at five independent compositions where the GAP model was never trained listed here as molar fraction of KCl (25%, 40%, 45%, 61% and 75%). At each of these compositions, RIM simulations are used to melt small unit cells of 64-66 atoms at 3000K, and are then equilibrated to 1200K. To ensure that the validation samples are well decorrelated the MD is continue at 1200K for 13.5 ns. The first 12.5 ns are discarded

and twenty uniform snapshots are obtained from the last 1 ns of this run (i.e. 50ps apart to ensure there is no correlation between obtained configurations and are spread across long time scales). Single point DFT-SCAN calculations are performed on these 100 (i.e. 20 x 5 composition) to obtain atomic forces. We use the GAP model to compute forces and directly compare them against the DFT-derived values to assess if the model provides consistent accuracy as a test of transferability. For comparison, we also report the atomic forces computed by RIM and the state-of-the-art polarizable ion model (PIM) fitted to DFT reported by Ishii *et al.* [59]. The results are shown in Fig. 3. Notice that GAP provides near DFT-SCAN accuracy ($0.1\text{eV}/\text{\AA}$) across compositions. The RIM model has the largest prediction error. The PIM model a 6-9 $\times$ increase in error in predicting DFT-SCAN derived forces as compared to the GAP model. The higher systematic error in RIM and PIM can be attributed to the limitations imposed by their inflexible functional forms and parameters.

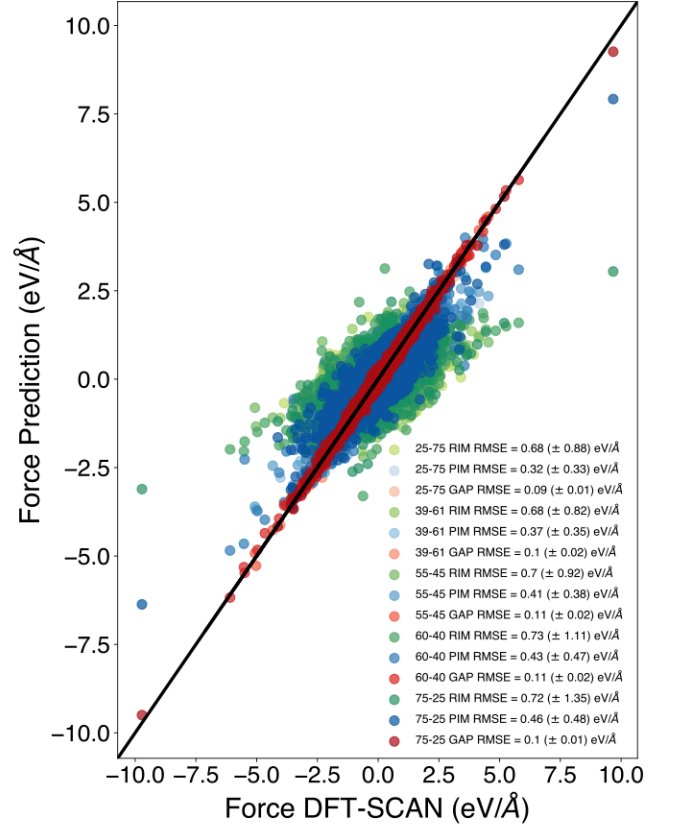


FIG. 3. Comparison of RIM, PIM, and GAP models against exact DFT-SCAN-derived forces for 100 diverse configurations across five independent compositions not explicitly included in the GAP training. The LiCl-KCl composition as %mol fractions for each validation group are reported in the legend. The standard deviation of the fit is reported in plot legend parentheses.

B. Comparison of composition, temperature dependent HEXRD measured structures with GAP-MD prediction

The final GAP model obtained from the workflow in Fig. 1 is used to perform simulated annealing with a system size over ~ 1000 atoms at each of the experimental compositions and temperatures (Table III) [60]. We performed GAP MD using the LAMMPS software package compiled with the QUIP pair style [61, 62]. The structure prediction with simulated annealing is performed in three steps: 1) Each simulation condition is initially thermalized at 1100K in the (NVT) ensemble [63, 64], followed by volume relaxation in an isothermal-isobaric (NPT) ensemble with a pressure coupling of 1 bar [65–67], 2) the temperature is increased to 1700K over 200 ps in the NPT ensemble, and 3) the system is cooled from 1700K to the target temperature over 200 ps of NPT simulations. All heating and cooling MD simulations used a time step of 0.5 fs. At the target temperature the volume of the system is relaxed for 2 ns with a time-step of 1.0 fs and the last 1 ns is used for computing the structure. Further information on the number of atoms, starting densities, and densities estimated from GAP-MD for each composition and temperature are listed in Table III (includes outside of experimental conditions).

To determine the quality of the generated GAP model, the structures obtained from simulations are validated by comparison to HEXRD measurements. The structure factors and PDFs from experiments and simulations are shown in Fig. 4. The corresponding snapshot of the simulation systems have been visualized in the Fig. 5. Note that such a comparison at temperatures just above the melting points has not been previously reported. The simulated PDF ($G(r)$) and structure factors ($S(q)$) exhibit excellent agreement with those obtained from HEXRD across all examined compositions. Specifically, GAP accurately predicts the structure of the eutectic (Fig. 4a, 4b) despite lacking any training data specifically in this composition regime. These results indicate that the GAP model is capable of predicting the structure of LiCl-KCl across all relevant compositions. This performance is a direct result of the use of AL to sample diverse configurations, and introduces the potential to develop ML-IP for binary MS mixtures of arbitrary composition.

Use of the validated GAP models provides insight into the atomistic ordering of LiCl-KCl melts across all compositions. As KCl content decreases in the mixture, the peak in the pair distribution curve near $3.1 \text{ \AA}$ also decreases, which can be attributed to the decreased amount of K-Cl bonding in the melt (Fig. 4a). This also manifests in the reduced structure factor, as the peaks near $4 \text{ \AA}^{-1}$ and $5 \text{ \AA}^{-1}$ both increase (Fig. 4b). Unlike MgCl_2 -KCl mixtures, which contain a single network former (Cl-Mg-Cl) and a single network breaker (K-Cl), the components in LiCl-KCl are both network breakers [60, 69]. Therefore, we do not expect that the variation of LiCl

and KCl content will change the mid-range (i.e., on the scale of 1 nm) ordering of the melt. To further elucidate the preferential interaction/exclusion of K/Li with Cl at short range (i.e., $r < 5 \text{ \AA}$), we have computed the local/bulk partition coefficient ($K_P(r)$) which is defined as

$$K_P = \frac{(\langle n_\alpha(r) \rangle / \langle n_\beta(r) \rangle)^{local}}{(n_\alpha^{tot} / n_\beta^{tot})^{bulk}} \quad (2)$$

where $\langle n_X(r) \rangle$ is the cumulative number distribution function of X ($X = \alpha$ for K around Cl, $X = \beta$ for Li around Cl) at a distance r and n_X^{tot} is the total number of X in the simulation box. $K_P > 1$ indicated a preferential interaction of K with respect to Cl [70, 71]. Equation 2 can be interpreted as $K_P(r) > 1$ in regions where K preferentially interacts with Cl and $K_P(r) < 1$ in regions where Li preferentially interacts with Cl. The composition dependence of $K_P(r)$ estimated at a fixed temperature of 747K is shown in Fig. 6. All three compositions exhibit $K_P(r) > 1$ at short range, showing a preferential interaction of K with Cl. Further, at short range the maximum values of $K_P(r)$ show a trend of 70:30 > Eutectic > 50:50, indicating an increase of preferential interaction of K with respect to Cl with decreasing content of KCl in the mixture. Complementary to the information in Fig. 4c, the temperature dependence of $K_P(r)$ at the 70:30 composition are visualized in Fig. 7. It can be observed that temperature does not lead to any significant change in $K_P(r)$. We also examined the coordination number of Li-Cl and K-Cl of LiCl-KCl mixtures with different compositions by integrating the partial PDF to the first minimum, and found that the coordination number decreases as the KCl content increases in the mixture at similar temperatures (i.e., 747K to 757K). This might be due to the larger size of K ions that reduce the number density of the melt. Increased temperature also leads to Li-Cl and K-Cl coordination numbers decreasing, however, the change is difficult to distinguish when the temperature variation magnitude is less than 100K. Details of the coordination number changes are listed in Table IV.

C. Thermal conductivity from GAP-MD

Finally, we assessed whether the accurate prediction of structure translated to the accurate prediction of thermophysical properties. Specifically, ML-IP trained at the level of DFT-SCAN has also shown to significantly improve the prediction of thermal conductivity relative to GGA [72]. Hence we compute the thermal conductivity of LiCl-KCl at the eutectic composition. We estimated the thermal conductivity by using the wave method [73]. A detailed discussion on the wave method is available in Appendix B.

The GAP MD simulation was performed for Eutectic system at 645.15K with a elongated system [73]. In order to avoid finite size effect, a very large $L_z \approx 1000 \text{ \AA}$ which translates to a system size of 8,320 atoms (1.558

Composition (LiCl:KCl)	Number of atoms	Temperature (K)	Initial density (g.cm^{-3})	Density from GAP-MD (g.cm^{-3})
70-30	1120	747.15	1.602	1.696 (± 0.012)
		772.15	1.602	1.692 (± 0.012)
		797.15	1.602	1.690 (± 0.012)
		822.15	1.602	1.685 (± 0.012)
Eutectic	1040	645.15	1.622	1.558 (± 0.015)
50-50	1024	757.15	1.584	1.555 (± 0.015)
30-70	1120	900	1.583	1.580 (± 0.012)
80-20	960	800	1.578	1.764 (± 0.012)

TABLE III. GAP MD simulation set up. A 50-50 composition of LiCl:KCl with 1024 atom system is equal to 256 anion-cation pairs for each salt. The initial density and target temperature used for MD simulation. The density computed from the last 1 ns of GAP-MD based NPT simulation along with standard deviations (parenthesis) are reported in the last column. The additional compositions 80-20, and 30-70 are computed to show case the capability of the GAP model to predict outside of experiment conditions.

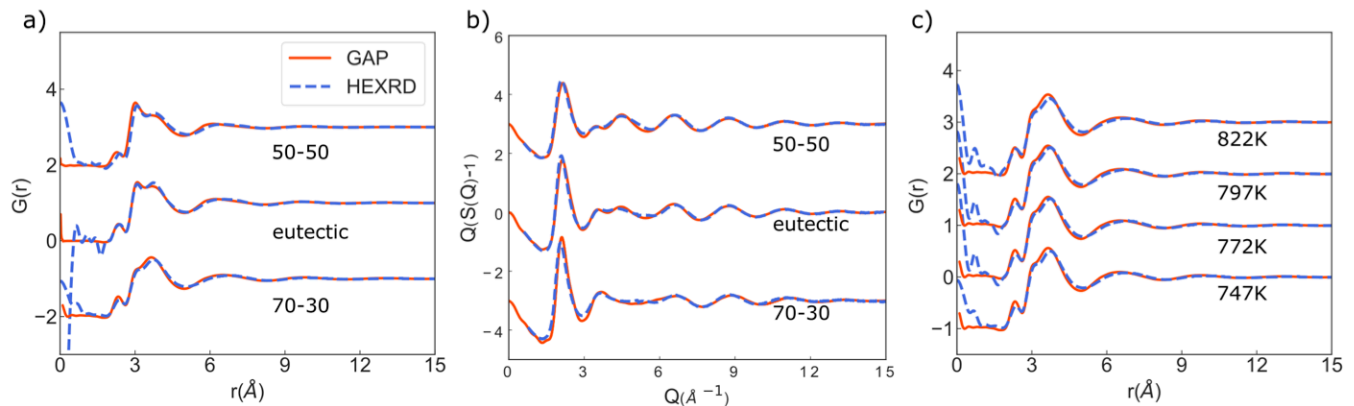


FIG. 4. GAP-predicted structures of molten LiCl-KCl compared to those measured by HEXRD: (a) PDFs and (b) reduced structure factors of LiCl-KCl with compositions of 50-50 mol%, 58.5-41.5 mol% (eutectic) and 70-30 mol%, at 757.15K, 645.15K, and 747.15K, respectively; (c) PDFs of molten 70-30 mol% LiCl-KCl at 747K, 772K, 797K, and 822K.

Composition	Temperature (K)	Li-Cl	K-Cl
70-30	747	4.42	7.28
	772	4.37	7.40
	797	4.45	7.27
	822	4.42	7.37
Eutectic	645	4.56	7.50
	747	4.21	6.90
50-50	757	4.06	6.87

TABLE IV. Coordination numbers estimated from GAP-MD

g. cm^{-3}) is chosen. For benchmark purposes, we also set up a simulation using RIM model with the same number of atoms and elongation (1.512 g. cm^{-3}). The simulations are performed with in NVT ensemble with a global stochastic velocity rescaling thermostat [74]. A time step of 0.5 fs is used and the simulations are performed for 500 ps. The results are reported in Table V. The value obtained from GAP MD (0.659 W/(m.K)) is very close to the experimental value (0.690 W/(m.K)) [44]. Encouraged by these results, we will undertake comprehensive composition dependent transport property studies in the future.

IV. CONCLUSIONS

In conclusion, here we report the development of a composition transferable GAP model for LiCl-KCl melts. We show that the model can effectively reproduce DFT-SCAN level forces within the limit of the training accuracy for five independent melt composition that was no included in the model training process. We also show that the GAP model can accurately predicted experimentally measured eutectic composition that was also never seen during the model training. Finally, the thermal conductivity computed using the GAP model for the eutectic composition at 645K shows close agreement with that obtained from experiment. The GAP model, training data, and the MD trajectories have been deposited in the Material Data facility (<https://materialsdatafacility.org/>) [75, 76] for public access [77].

The AL-enabled, and HEXRD-validated, ML-IP workflow described in this study allows for the rapid characterization of multi-component molten salts of arbitrary compositions and will find broad applicability to other eutectic salt systems. This workflow accelerates the develop-

Method	c_v (J/(K.mol))	γ	c_s (Å/ps)	D_T (Å ² /ps)	λ (W/(m.K))
GAP	34.1	1.21	0.434	27.5(2)	0.659
RIM	29.8	1.20	0.355	43.9(2)	0.856
Experiment [44]	-	-	-	-	0.690

TABLE V. Fit parameters for eqn. B3 and the estimated thermal conductivity. Linear fit $D_T(k)$ is performed as function of k for all the points where statistical error up to 10^{-2} (paranthesis) and extrapolated to $k=0$ as explained in Ref [73].

ment of DFT accurate composition transferable ML-IP which can be used for high accuracy property prediction across various compositions and simulation conditions, while simultaneously enabling the real-time deconvolution of complex chemical structures obtained from HEXRD experiments.

ACKNOWLEDGMENTS

This material is based upon work supported by Laboratory Directed Research and Development (LDRD-2020-0226, LDRD-CLS-1-630) funding from Argonne National Laboratory, provided by the Director, Office of Science, of the U.S. Department of Energy under Contract No. DE-AC02-06CH11357. This research was supported by ExaLearn Co-design Center of the Exascale Computing Project (17-SC-20-SC), a collaborative effort of the U.S. Department of Energy Office of Science and the National Nuclear Security Administration [78]. We gratefully acknowledge the computing resources provided on Bebop; a high-performance computing cluster operated by the Laboratory Computing Resource Center at Argonne National Laboratory. This research used resources of the Argonne Leadership Computing Facility, a DOE Office of Science User Facility supported under Contract DE-AC02-06CH11357. HEXRD measurements were made on beamline 6-ID-D at the Advanced Photon Source, a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under Contract No. DE-AC02-06CH11357. Argonne National Laboratory’s work was supported by the U.S. Department of Energy, Office of Science, under contract DE-AC02-06CH11357. G. S. would like to thank Prof. Gabor Csányi for constructive feedback on the manuscript and fruitful discussions on GAP model fitting. G.S. would like to Dr. Anand Narayanan Krishnamoorthy for fruitful discussions on local/bulk partition coefficient.

Appendix A: Active Learning

The pseudocode for the clustering based active learning [47], described below, bears overlap with the previous work of Sivaraman *et al.* [42, 48]. For each composition of LiCl-KCl melt:

Initialization: trajectory, distance measure, target accuracy.

1. Configurations of the input trajectory are featurized by using distance matrices.
2. Perform unsupervised clustering based on HDBSCAN algorithm [79] to obtain uncorrelated clusters. The distance measure of root mean square deviation of atomic positions is used.
3. Training and test configurations are sequentially drawn from the clustering performed on large pool of RIM simulation configuration.
4. Perform single point DFT configuration
5. Perform hyperparameter tuning using Bayesian optimization (BO) as implemented in the GPyOpt library [80] to get the best GAP model for a chosen set of training configurations as validated against independent test configurations.
6. Draw more samples from the clustered configurations if the GAP target accuracy of ≤ 5 meV/atom has not been achieved and repeat [3-5].
7. Exit if the target accuracy is achieved. For the accuracy we use Mean Absolute Error in GAP predicted energy with respect to DFT in the unit of meV/atom.

The number of training configurations drawn at each iteration depends on the outcome of unsupervised clustering, and hence adjusts automatically with the input chemical system of interest [47, 81]. A step-by-step tutorial is also available elsewhere [82].

Appendix B: Thermal conductivity

The thermal conductivity estimated from virial stress based heat flux implemented with LAMMPS is not valid beyond pair interaction models [61, 83]. Hence they cannot be applied to many-body ML-IP such GAP model. To bypass this problem we will estimate the thermal diffusivity of eutectic liquid using the thermally driven liquid density fluctuation at hydrodynamic limit (i.e. as system size goes to infinity, $k \rightarrow 0$) [73, 84]. The most relevant details of the approach are reproduced below. The density field of the liquid in bulk limit ($\rho(r, t)$) is computed from equilibrium MD simulation. For a periodic system with dimension of $\{L_x, L_y, L_z\}$, the Fourier transform of the density field $\rho(r, t)$ in a bulk liquid is defined as:

$$\tilde{\rho}(k, t) = \frac{1}{V} \sum_{i=1}^N \exp(-ik \cdot r_i(t)), \quad (B1)$$

where the reciprocal lattice vector, $k = \{\frac{2\pi n_x}{L_x}, \frac{2\pi n_y}{L_y}, \frac{2\pi n_z}{L_z}\}$. The power spectrum of the density field fluctuation with frequency ω is defined as:

$$S(k, \omega) = \frac{1}{T_0} \int_0^{T_0} \langle \tilde{\rho}(k, 0) \tilde{\rho}(k, t) \rangle \exp(-i\omega t) dt \quad (B2)$$

The $S(k, \omega)$ derived from the MD simulation can be fitted to power spectrum equation derived from hydrodynamic theory:

$$S(k, \omega) = \frac{S(k)}{2\pi} \left[\left(\frac{\gamma - 1}{\gamma} \right) \frac{2D_T k^2}{\omega^2 + (D_T k^2)^2} + \frac{1}{\gamma} \left(\frac{\Gamma k^2}{(\omega + c_s k)^2 + (\Gamma k^2)^2} + \frac{\Gamma k^2}{(\omega - c_s k)^2 + (\Gamma k^2)^2} \right) \right] \quad (B3)$$

where, $\gamma = c_P/c_V$ is the ratio of particle specific heat capacities (defined below), D_T is the thermal diffusivity, c_s is the adiabatic speed of sound, b is the kinematic longitudinal viscosity, and the sound attenuation constant $\Gamma = (\gamma - 1)D_T/2 + b/2$. The first term in eqn. B3 is related to thermal transport and the last two terms corresponds to sound wave travelling in opposite direction. Further details on the methodology along with the code is discussed in detail elsewhere [73].

Now we will briefly discuss the thermodynamic relation relevant to quantities that arise in Eqn. B3. These quantities are estimated from numeric partial derivatives with respect to state variables as defined below:

$$c_V = \frac{1}{N} \left(\frac{\partial E}{\partial T} \right)_V \quad (B4)$$

$$c_P - c_V = -\frac{T}{N} \left(\frac{\partial P}{\partial T} \right)_V^2 \left(\frac{\partial P}{\partial V} \right)_T^{-1} \quad (B5)$$

$$c_s^2 = -\frac{V^2 \gamma}{Nm} \left(\frac{\partial P}{\partial V} \right)_T \quad (B6)$$

As described in Ref. [73], D_T and b will be treated as the fit parameters. Once the D_T is obtained by a linear fit, the thermal conductivity (λ) can be computed using the following relation:

$$\lambda = \rho_0 c_P D_T \quad (B7)$$

where ρ_0 is the liquid density.

-
- [1] J. Dupont, From molten salts to ionic liquids: a “nano” journey, *Accounts of Chemical Research* **44**, 1223 (2011).
 - [2] H. Kim, D. A. Boysen, J. M. Newhouse, B. L. Spatocco, B. Chung, P. J. Burke, D. J. Bradwell, K. Jiang, A. A. Tomaszowska, K. Wang, *et al.*, Liquid metal batteries: past, present, and future, *Chemical Reviews* **113**, 2075 (2013).
 - [3] H. Gougar, D. Petti, P. Demkowicz, W. Windes, G. Strydom, J. Kinsey, J. Ortensi, M. Plummer, W. Skerjanc, R. Williamson, *et al.*, The us department of energy’s high temperature reactor research and development program—progress as of 2019, *Nuclear Engineering and Design* **358**, 110397 (2020).
 - [4] J. Guo, N. Hoyt, and M. Williamson, Multielectrode array sensors to enable long-duration corrosion monitoring and control of concentrating solar power systems, *Journal of Electroanalytical Chemistry* **884**, 115064 (2021).
 - [5] M. Mehos, C. Turchi, J. Vidal, M. Wagner, Z. Ma, C. Ho, W. Kolb, C. Andraka, and A. Kruizenga, *Concentrating solar power Gen3 demonstration roadmap*, Tech. Rep. (National Renewable Energy Lab, Golden, CO, United States, 2017).
 - [6] J. Jerden, *Molten Salt Thermophysical Properties Database Development: 2019 Update*, Tech. Rep. (Argonne National Lab, Argonne, IL, United States, 2019).
 - [7] J. W. McMurray, K. Johnson, C. Agca, B. R. Betzler, D. J. Kropaczek, T. M. Besmann, D. Andersson, and N. Ezell, *Roadmap for thermal property measurements of Molten Salt Reactor systems*, Tech. Rep. (Oak Ridge National Lab.(ORNL), Oak Ridge, TN (United States), 2021).
 - [8] S. Jayaraman, A. P. Thompson, and O. A. von Lilienfeld, Molten salt eutectics from atomistic simulations, *Physical Review E* **84**, 030201 (2011).
 - [9] M. Born and J. E. Mayer, Zur gittertheorie der ionenkristalle, *Zeitschrift für Physik* **75**, 1 (1932).

- [10] F. Fumi and M. Tosi, Ionic sizes and born repulsive parameters in the NaCl-type alkali halides—I: The Huggins-Mayer and Pauling forms, *Journal of Physics and Chemistry of Solids* **25**, 31 (1964).
- [11] P. A. Madden and M. Wilson, ‘covalent’ effects in ‘ionic’ systems, *Chemical Society Reviews* **25**, 339 (1996).
- [12] M. M. Ghahremanpour, P. J. Van Maaren, and D. Van Der Spoel, The alexandria library, a quantum-chemical database of molecular properties for force field development, *Scientific Data* **5**, 1 (2018).
- [13] M. Salanne and P. A. Madden, Polarization effects in ionic solids and melts, *Molecular Physics* **109**, 2299 (2011).
- [14] H. Wang, R. S. DeFever, Y. Zhang, F. Wu, S. Roy, V. S. Bryantsev, C. J. Margulis, and E. J. Maginn, Comparison of fixed charge and polarizable models for predicting the structural, thermodynamic, and transport properties of molten alkali chlorides, *The Journal of Chemical Physics* **153**, 214502 (2020).
- [15] J. Behler and G. Csányi, Machine learning potentials for extended systems: a perspective, *The European Physical Journal B* **94**, 1 (2021).
- [16] V. L. Deringer, A. P. Bartók, N. Bernstein, D. M. Wilkins, M. Ceriotti, and G. Csányi, Gaussian process regression for materials and molecules, *Chemical Reviews* **121**, 10073 (2021).
- [17] M. Ceriotti, C. Clementi, and O. Anatole von Lilienfeld, Machine learning meets chemical physics, *The Journal of Chemical Physics* **154**, 160401 (2021).
- [18] Y. LeCun, Y. Bengio, and G. Hinton, Deep learning, *nature* **521**, 436 (2015).
- [19] A. P. Bartók, M. C. Payne, R. Kondor, and G. Csányi, Gaussian approximation potentials: The accuracy of quantum mechanics, without the electrons, *Physical review letters* **104**, 136403 (2010).
- [20] J. Behler and M. Parrinello, Generalized neural-network representation of high-dimensional potential-energy surfaces, *Physical review letters* **98**, 146401 (2007).
- [21] I. S. Novikov, K. Gubaev, E. V. Podryabinkin, and A. V. Shapeev, The mlip package: moment tensor potentials with mpi and active learning, *Machine Learning: Science and Technology* **2**, 025002 (2020).
- [22] L. Zhang, J. Han, H. Wang, R. Car, and E. Weinan, Deep potential molecular dynamics: a scalable model with the accuracy of quantum mechanics, *Physical review letters* **120**, 143001 (2018).
- [23] A. M. Cooper, J. Kästner, A. Urban, and N. Artrith, Efficient training of ann potentials by including atomic forces via taylor expansion and application to water and a transition-metal oxide, *npj Computational Materials* **6**, 1 (2020).
- [24] A. P. Thompson, L. P. Swiler, C. R. Trott, S. M. Foiles, and G. J. Tucker, Spectral neighbor analysis method for automated generation of quantum-accurate interatomic potentials, *Journal of Computational Physics* **285**, 316 (2015).
- [25] S. Batzner, A. Musaelian, L. Sun, M. Geiger, J. P. Mailoa, M. Kornbluth, N. Molinari, T. E. Smidt, and B. Kozinsky, Se (3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials, *arXiv preprint arXiv:2101.03164* (2021).
- [26] F. Musil, M. Veit, A. Goscinski, G. Fraux, M. J. Willatt, M. Stricker, T. Junge, and M. Ceriotti, Efficient implementation of atom-density representations, *The Journal of Chemical Physics* **154**, 114109 (2021).
- [27] J. S. Smith, O. Isayev, and A. E. Roitberg, Ani-1: an extensible neural network potential with dft accuracy at force field computational cost, *Chemical science* **8**, 3192 (2017).
- [28] K. T. Schütt, H. E. Sauceda, P.-J. Kindermans, A. Tkatchenko, and K.-R. Müller, Schnet—a deep learning architecture for molecules and materials, *The Journal of Chemical Physics* **148**, 241722 (2018).
- [29] D. P. Kovács, C. v. d. Oord, J. Kucera, A. E. Allen, D. J. Cole, C. Ortner, and G. Csányi, Linear atomic cluster expansion force fields for organic molecules: beyond rmse, *Journal of chemical theory and computation* **17**, 7696 (2021).
- [30] D. M. d. O. Zapiain, M. A. Wood, N. Lubbers, C. Z. Pereyra, A. P. Thompson, and D. Perez, Training data selection for accuracy and transferability of interatomic potentials, *arXiv preprint arXiv:2201.09829* (2022).
- [31] S. Tovey, A. Narayanan Krishnamoorthy, G. Sivaraman, J. Guo, C. Benmore, A. Heuer, and C. Holm, Dft accurate interatomic potential for molten nacl from machine learning, *The Journal of Physical Chemistry C* **124**, 25760 (2020).
- [32] S.-C. Lee, Y. Zhai, Z. Li, N. P. Walter, M. Rose, B. J. Heuser, and Y. Z., Comparative studies of the structural and transport properties of molten salt fli-nak using the machine-learned neural network and reparametrized classical forcefields, *The Journal of Physical Chemistry B* **125**, 10562 (2021), pMID: 34496565, <https://doi.org/10.1021/acs.jpcc.1c05608>.
- [33] S. T. Lam, Q.-J. Li, R. Ballinger, C. Forsberg, and J. Li, Modeling lif and flibe molten salts with robust neural network interatomic potential, *ACS Applied Materials & Interfaces* (2021).
- [34] M.-T. Nguyen, R. Rousseau, P. D. Paviot, and V.-A. Glezakou, Actinide molten salts: A machine-learning potential molecular dynamics study, *ACS Applied Materials & Interfaces* (2021).
- [35] J. Zhang, J. Fuller, and Q. An, Coordination and thermophysical properties of transition metal chlorocomplexes in lici-kcl eutectic, *The Journal of Physical Chemistry B* **125**, 8876 (2021).
- [36] Q.-J. Li, E. Küçükbenli, S. Lam, B. Khaykovich, E. Kaxiras, and J. Li, Development of robust neural-network interatomic potential for molten salt, *Cell Reports Physical Science* **2**, 100359 (2021).
- [37] K. Liu, Y.-L. Liu, Z.-F. Chai, and W.-Q. Shi, Evaluation of the electroextractions of ce and nd from lici-kcl molten salt using liquid ga electrode, *Journal of The Electrochemical Society* **164**, D169 (2017).
- [38] P. D. Myers Jr and D. Y. Goswami, Thermal energy storage using chloride salts and their eutectics, *Applied Thermal Engineering* **109**, 889 (2016).
- [39] A. Merwin, M. A. Williamson, J. L. Willit, and D. Chidambaram, Metallic lithium and the reduction of actinide oxides, *Journal of the Electrochemical Society* **164**, H5236 (2017).
- [40] A. P. Bartók, R. Kondor, and G. Csányi, On representing chemical environments, *Physical Review B* **87**, 184115 (2013).
- [41] V. L. Deringer and G. Csányi, Machine learning based interatomic potential for amorphous carbon, *Physical Review B* **95**, 094203 (2017).
- [42] G. Sivaraman, J. Guo, L. Ward, N. Hoyt, M. Williamson,

- I. Foster, C. Benmore, and N. Jackson, Automated development of molten salt machine learning potentials: Application to LiCl , *The Journal of Physical Chemistry Letters* **12**, 4278 (2021).
- [43] C. G. Staacke, H. H. Heenen, C. Scheurer, G. Csányi, K. Reuter, and J. T. Margraf, On the role of long-range electrostatics in machine-learned interatomic potentials for complex battery materials, *ACS Applied Energy Materials* **4**, 12562 (2021).
- [44] G. J. Janz, *Molten salts handbook* (Elsevier, 2013).
- [45] L. Ward, G. Sivaraman, J. G. Pauloski, Y. Babuji, R. Chard, N. Dandu, P. C. Redfern, R. S. Assary, K. Chard, L. A. Curtiss, *et al.*, Colmena: Scalable machine-learning-based steering of ensemble simulations for high performance computing, arXiv preprint arXiv:2110.02827 (2021).
- [46] A. Laio and M. Parrinello, Escaping free-energy minima, *Proceedings of the National Academy of Sciences* **99**, 12562 (2002).
- [47] G. Sivaraman, A. N. Krishnamoorthy, M. Baur, C. Holm, M. Stan, G. Csányi, C. Benmore, and Á. Vázquez-Mayagoitia, Machine-learned interatomic potentials by active learning: amorphous and liquid hafnium dioxide, *npj Computational Materials* **6**, 1 (2020).
- [48] G. Sivaraman, L. Gallington, A. N. Krishnamoorthy, M. Stan, G. Csányi, Á. Vázquez-Mayagoitia, and C. J. Benmore, Experimentally driven automated machine-learned interatomic potential for a refractory oxide, *Physical Review Letters* **126**, 156002 (2021).
- [49] J. Sun, A. Ruzsinszky, and J. P. Perdew, Strongly constrained and appropriately normed semilocal density functional, *Physical review letters* **115**, 036402 (2015).
- [50] J. Sun, R. C. Remsing, Y. Zhang, Z. Sun, A. Ruzsinszky, H. Peng, Z. Yang, A. Paul, U. Waghmare, X. Wu, *et al.*, Accurate first-principles structures and energies of diversely bonded systems from an efficient density functional, *Nature chemistry* **8**, 831 (2016).
- [51] L. C. Erhard, J. Rohrer, K. Albe, and V. L. Deringer, A machine-learned interatomic potential for silica and its relation to empirical models, *npj Computational Materials* **8**, 1 (2022).
- [52] G. Kresse and J. Furthmüller, Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set, *Physical review B* **54**, 11169 (1996).
- [53] P. E. Blöchl, Projector augmented-wave method, *Physical review B* **50**, 17953 (1994).
- [54] N. Bernstein, G. Csányi, and V. L. Deringer, De novo exploration and self-guided learning of potential-energy surfaces, *npj Computational Materials* **5**, 1 (2019).
- [55] L. Bonati and M. Parrinello, Silicon liquid structure and crystal nucleation from ab initio deep metadynamics, *Physical review letters* **121**, 265701 (2018).
- [56] M. Bonomi, Promoting transparency and reproducibility in enhanced molecular simulations, *Nature methods* **16**, 670 (2019).
- [57] A. Barducci, G. Bussi, and M. Parrinello, Well-tempered metadynamics: a smoothly converging and tunable free-energy method, *Physical review letters* **100**, 020603 (2008).
- [58] G. A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni, and G. Bussi, Plumed 2: New feathers for an old bird, *Computer Physics Communications* **185**, 604 (2014).
- [59] Y. Ishii, S. Kasai, M. Salanne, and N. Ohtori, Transport coefficients and the stokes–einstein relation in molten alkali halides with polarisable ion model, *Molecular Physics* **113**, 2442 (2015).
- [60] F. Wu, S. Sharma, S. Roy, P. Halstenberg, L. C. Gallington, S. M. Mahurin, S. Dai, V. S. Bryantsev, A. S. Ivanov, and C. J. Margulis, Temperature dependence of short and intermediate range order in molten MgCl_2 and its mixture with KCl , *The Journal of Physical Chemistry B* **124**, 2892 (2020).
- [61] S. Plimpton, Fast parallel algorithms for short-range molecular dynamics, *Journal of computational physics* **117**, 1 (1995).
- [62] G. Csányi, S. Winfield, J. R. Kermode, A. De Vita, A. Comisso, N. Bernstein, and M. C. Payne, Expressive programming for computational physics in fortran 95+, *IoP Comput. Phys. Newsletter*, Spring 2007 (2007).
- [63] S. Nosé, A unified formulation of the constant temperature molecular dynamics methods, *The Journal of chemical physics* **81**, 511 (1984).
- [64] W. G. Hoover, Canonical dynamics: Equilibrium phase-space distributions, *Physical review A* **31**, 1695 (1985).
- [65] M. Parrinello and A. Rahman, Polymorphic transitions in single crystals: A new molecular dynamics method, *Journal of Applied physics* **52**, 7182 (1981).
- [66] G. J. Martyna, D. J. Tobias, and M. L. Klein, Constant pressure molecular dynamics algorithms, *The Journal of chemical physics* **101**, 4177 (1994).
- [67] W. Shinoda, M. Shiga, and M. Mikami, Rapid estimation of elastic constants by molecular dynamics simulation under constant stress, *Physical Review B* **69**, 134103 (2004).
- [68] K. Momma and F. Izumi, Vesta 3 for three-dimensional visualization of crystal, volumetric and morphology data, *Journal of applied crystallography* **44**, 1272 (2011).
- [69] M. Wilson and P. Madden, “prepeaks” and “first sharp diffraction peaks” in computer simulations of strong and fragile ionic liquids, *Physical review letters* **72**, 3033 (1994).
- [70] F. Rodriguez-Ropero and N. F. van der Vegt, Direct osmolyte–macromolecule interactions confer entropic stability to folded states, *The Journal of Physical Chemistry B* **118**, 7327 (2014).
- [71] A. N. Krishnamoorthy, J. Zeman, C. Holm, and J. Smiatek, Preferential solvation and ion association properties in aqueous dimethyl sulfoxide solutions, *Physical Chemistry Chemical Physics* **18**, 31312 (2016).
- [72] D. Tisi, L. Zhang, R. Bertossa, H. Wang, R. Car, and S. Baroni, Heat transport in liquid water from first-principles and deep neural network simulations, *Physical Review B* **104**, 224202 (2021).
- [73] B. Cheng and D. Frenkel, Computing the heat conductivity of fluids from density fluctuations, *Physical Review Letters* **125**, 130602 (2020).
- [74] G. Bussi, D. Donadio, and M. Parrinello, Canonical sampling through velocity rescaling, *The Journal of chemical physics* **126**, 014101 (2007).
- [75] B. Blaiszik, K. Chard, J. Pruyne, R. Ananthakrishnan, S. Tuecke, and I. Foster, The materials data facility: data services to advance materials science research, *Jom* **68**, 2045 (2016).
- [76] B. Blaiszik, L. Ward, M. Schwarting, J. Gaff, R. Chard, D. Pike, K. Chard, and I. Foster, A data ecosystem to support machine learning in materials science, *MRS Communications* **9**, 1125 (2019).

- [77] J. Guo, L. Ward, Y. Babuji, N. Hoyt, M. Williamson, I. Foster, N. Jackson, C. B. Benmore, and G. Sivaraman, Gap-ml model, training and md dataset for ‘a composition-transferable machine learning potential for lici-kcl molten salts validated by hexrd.’ (2022).
- [78] F. J. Alexander, J. Ang, J. A. Bilbrey, J. Balewski, T. Casey, R. Chard, J. Choi, S. Choudhury, B. Debusschere, A. M. DeGennaro, *et al.*, Co-design center for exascale machine learning technologies (exalearn), The International Journal of High Performance Computing Applications , 10943420211029302 (2020).
- [79] L. McInnes, J. Healy, and S. Astels, HDBSCAN: Hierarchical density based clustering, Journal of Open Source Software **2**, 205 (2017).
- [80] J. González and Z. Dai, GPyOpt: A Bayesian optimization framework in Python (2016).
- [81] G. Sivaraman, G. Csanyi, A. Vazquez-Mayagoitia, I. T. Foster, S. K. Wilke, R. Weber, and C. J. Benmore, A combined machine learning and high-energy x-ray diffraction approach to understanding liquid and amorphous metal oxides, Journal of the Physical Society of Japan **91**, 091009 (2022).
- [82] G. Sivaraman, Gap-al tutorial for the psi-k ml-ip 2021 tutorial workshop (2021).
- [83] P. Boone, H. Babaei, and C. E. Wilmer, Heat flux for many-body interactions: corrections to lammmps, Journal of chemical theory and computation **15**, 5579 (2019).
- [84] R. D. Mountain, Spectral distribution of scattered light in a simple fluid, Reviews of Modern Physics **38**, 205 (1966).

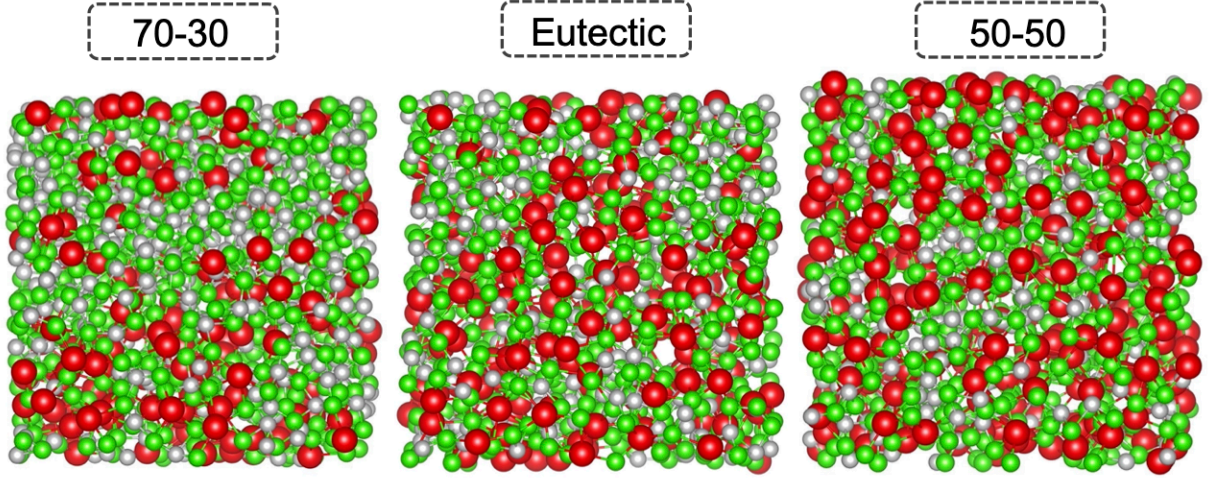


FIG. 5. Visualization of atomic configurations for LiCl-KCl melts with compositions of 50-50 mol%, 58.5-41.5 mol% (eutectic) and 70-30 mol%, at 757.15K, 645.15K, and 747.15K. The red, green and silver spheres each corresponds to atomic species for K, Cl and Li, respectively. The visualization is performed using VESTA software [68].

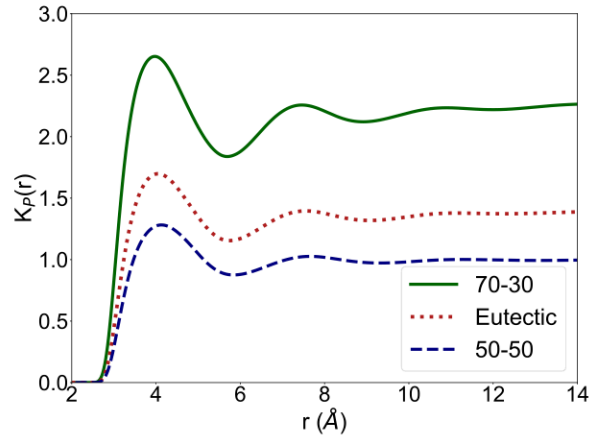


FIG. 6. Local/bulk partition coefficient computed at 747K as function of LiCl:KCl composition.

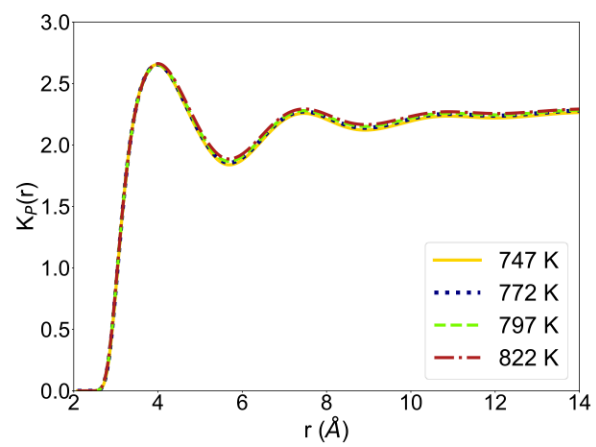


FIG. 7. Local/bulk partition coefficient computed for LiCl:KCl composition of 70:30 as function of temperature.